Knut Dundas Morå
fysikk@dundasmora.no, he/him

School of Underground Physics at Bertinoro

Statistics and Inference

for rare event searches

- *What is a statistical model?*
- *Does it describe your data?*
- *What kinds of conclusions can we draw?*

24

Summary of first topic

- We model our observations with a statistical model, usually in terms of probability distributions.

- We choose test statistics that distil the information we wish to learn from the data

- and often formulate questions in terms of hypothesis tests— given the data, should we favour one or the other?

- A particularly important hypothesis test is whether your data agrees with the distribution you use!

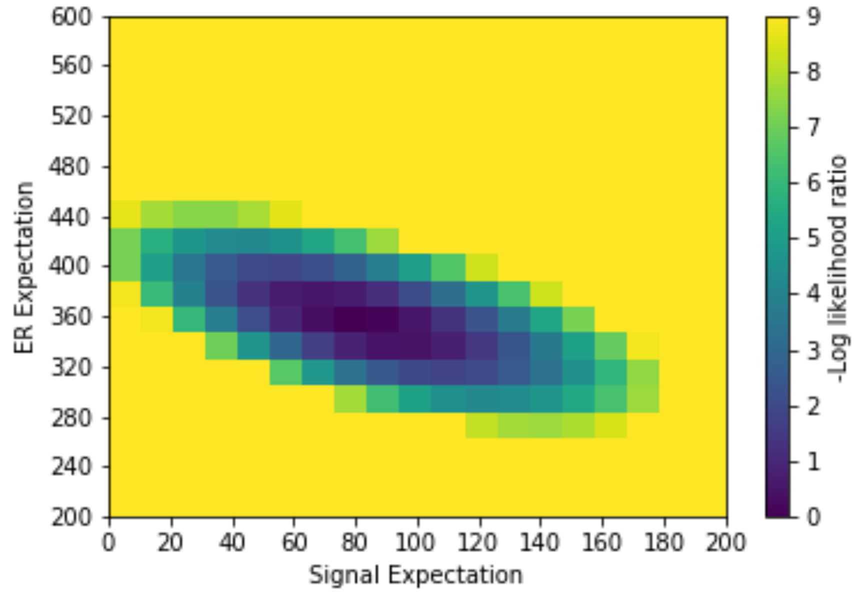
For today

- Example analyses
- Profile Likelihood
- Asymptotic distributions
- Look-Elsewhere effect

Profiling Likelihoods & nuisance parameters

SoUP 24

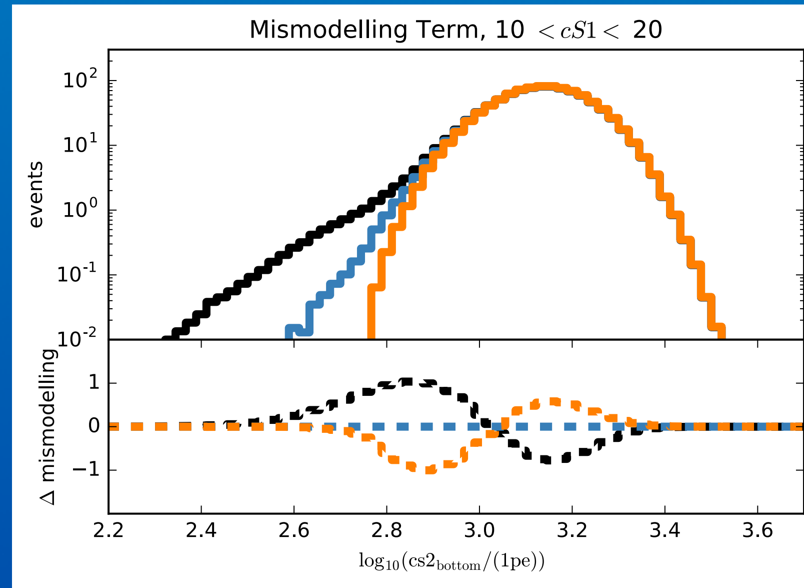
- We seldom have completely specified hypotheses
- Our background and signal models have uncertainties, parameterised by nuisance parameters (θ)— you'll see some examples in the next slides.
- The global best fit we denote with $\hat{s}, \hat{\theta}$
- However, we also want to test other s — for example s=0 for discovery significance or a range of s for confidence intervals.
- In these cases, we set the other nuisance parameters to their conditional best-fit $\hat{\hat{\theta}}$.



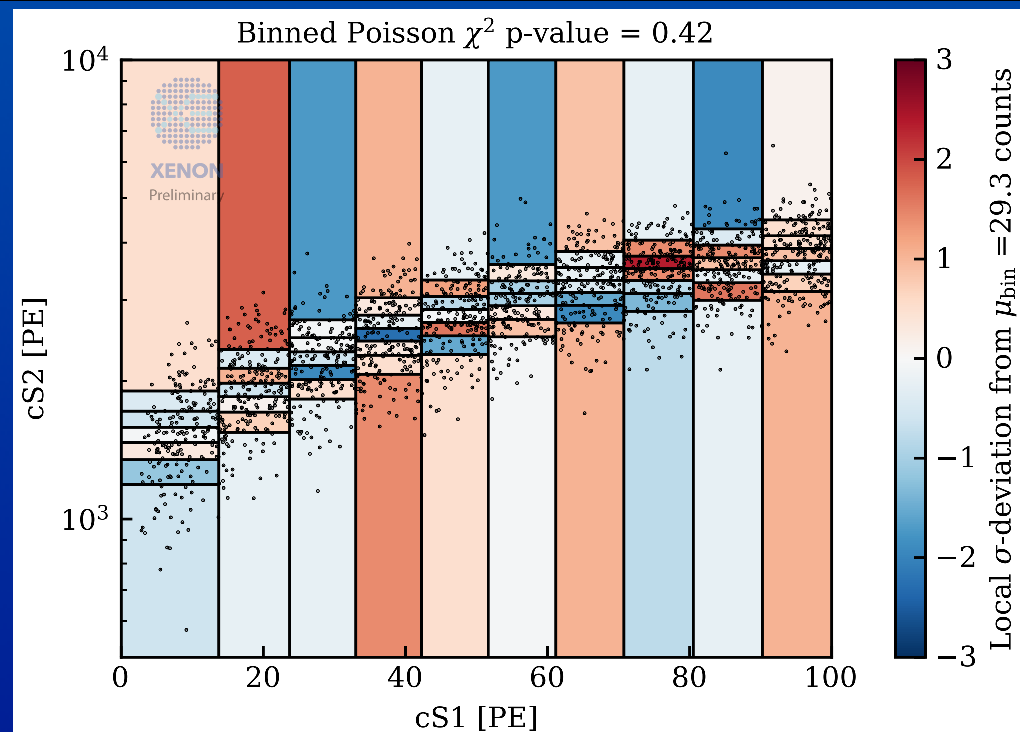
$$\delta \log \mathscr{L}(s, \hat{\hat{\boldsymbol{\theta}}})/\delta\theta_j = 0;$$

The likelihood relies on the model

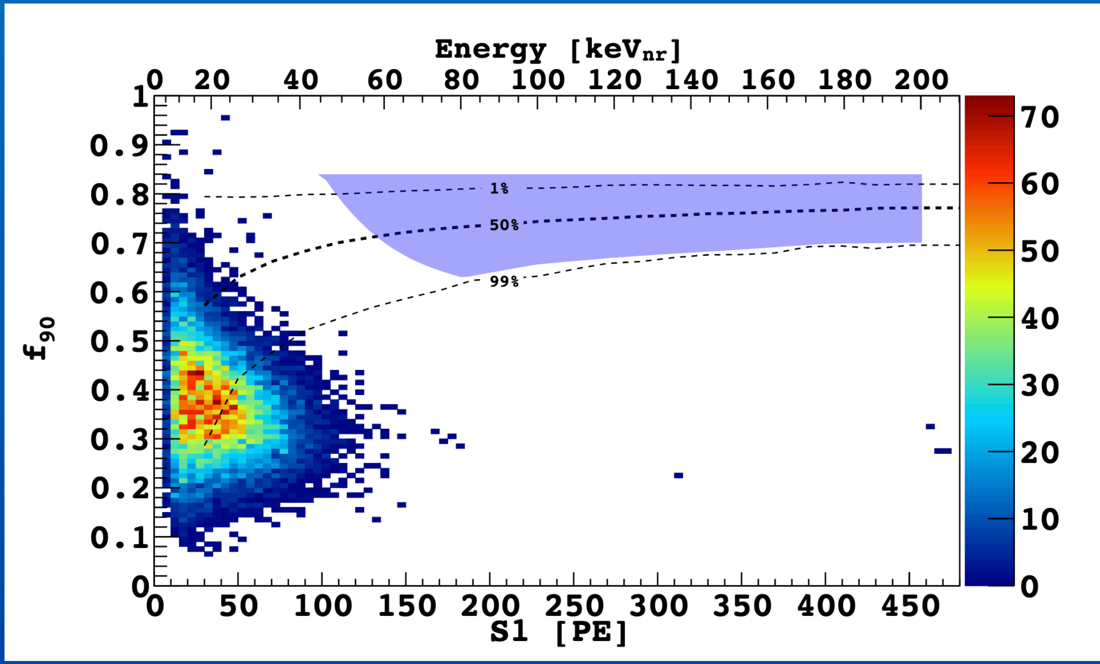
- The validity of the inference relies on the underlying model
- The signal model may be quite forgiving— if an excess is 10-20 events, far tails are less significant
- Experiments typically include uncertainties on background rates, but not always on the distribution used.
- XENON1T added a "signal-like" background shape to its ER background model to lower the chance of overconstraining the model.
- For XENONnT, this was replaced by a more careful selection of nuisance parameter directions, and a stronger focus on pre-defined goodness-of-fit tests chosen for their power to discover mismodelling



N. Priel et al. A model independent safeguard against background mismodeling for statistical inference. 2017(05):013–013, may 2017. doi: 10.1088/1475-7516/2017/05/013.



Counting Experiments

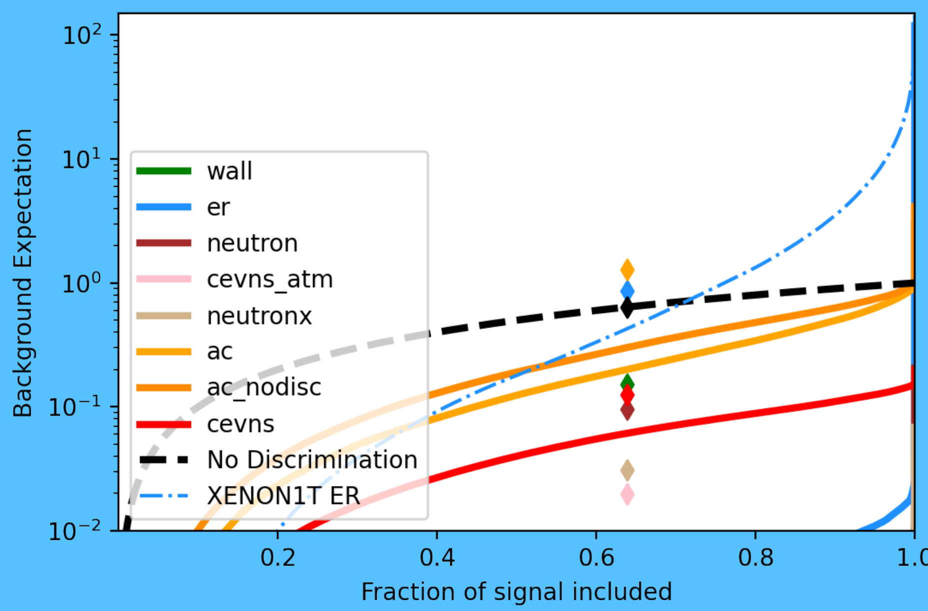


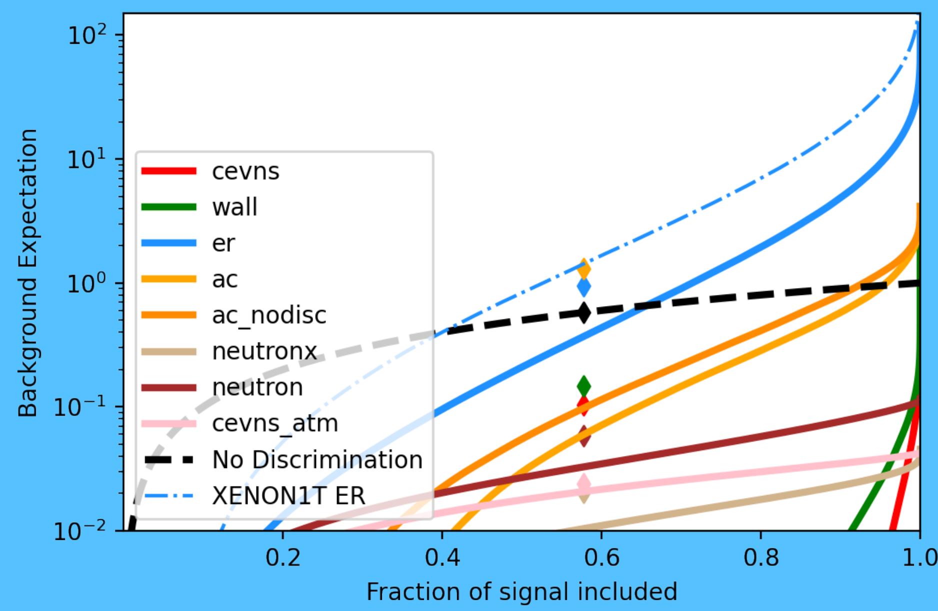
DarkSide-50 532-day https://arxiv.org/pdf/1802.07198

- "just" counting events— but the estimate of the background rate and acceptance can be as complicated as anything
- If there is no signal/background overlap *or* complete overlap, this may be the optimal sensitivity
- Otherwise, it might still be a worthwhile compromise if you're worried about whether you can model your background correctly

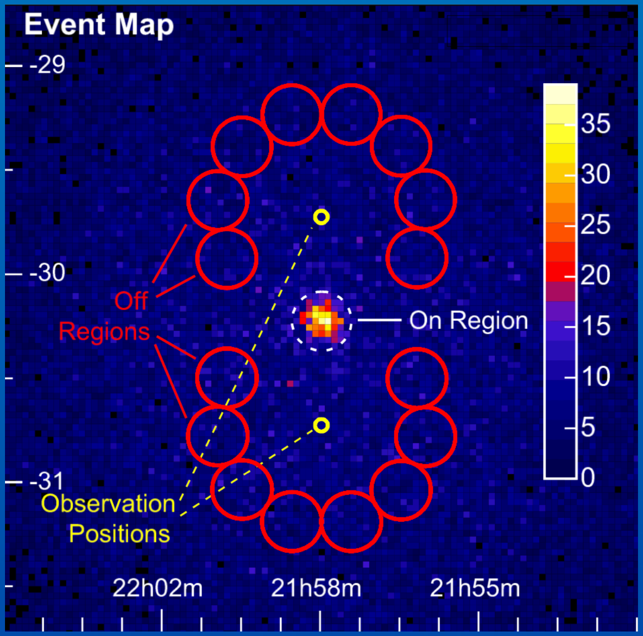
$$\mathscr{L}_{\rm sci}(s, \overrightarrow{\theta_s}, \vec{\theta}_b) = \mathrm{Poisson}(N_{\rm sci} \,|\, \mu_b(\vec{\theta}_b) + \mu_s(s, \overrightarrow{\theta_s}, \vec{\theta}_b))$$

However, shapes often matter





On-Off likelihoods

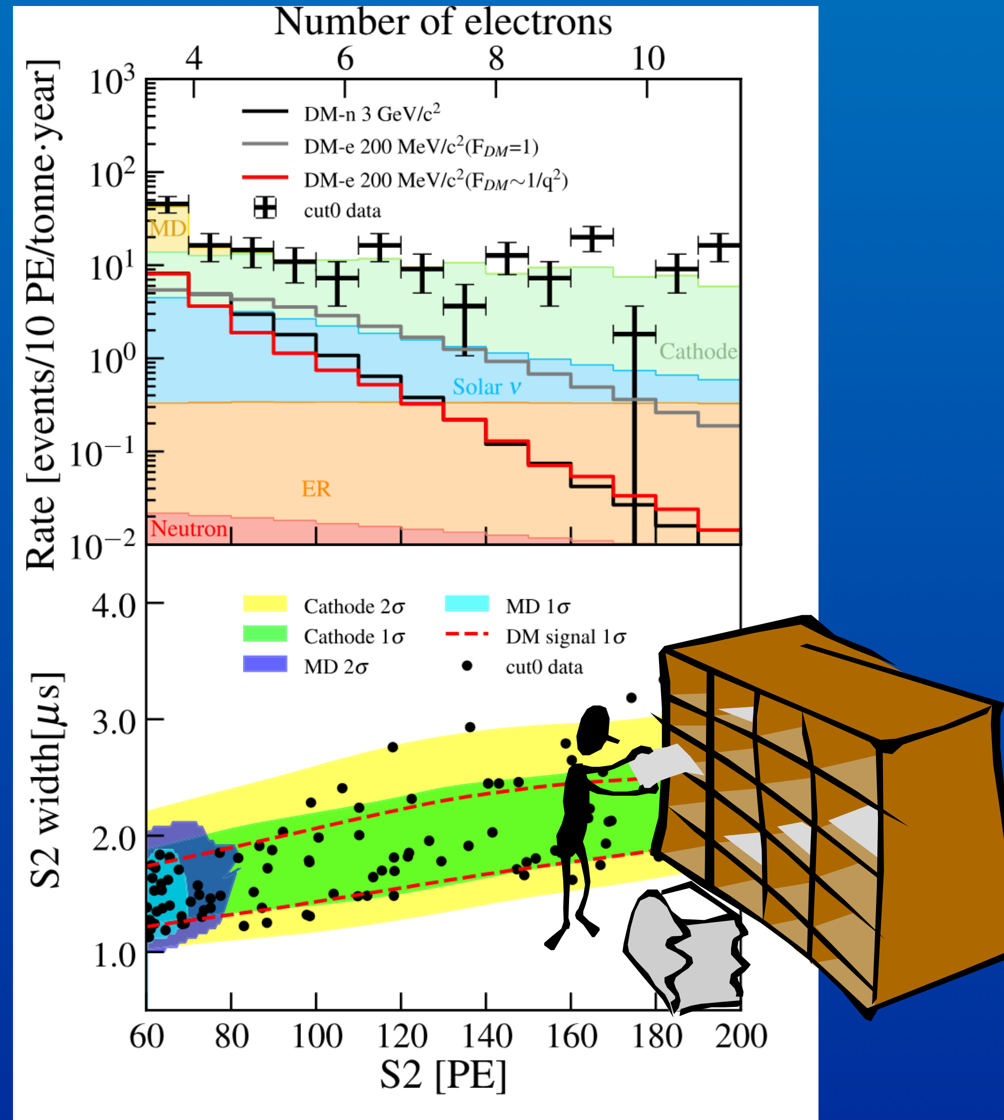


- WIMP searches rarely get to turn off their signal completely
- Directional dark matter searches and some axion searches, on the other hand can take representative data in a no/low signal and high signal state
- Also common in indirect detection

$$\mathcal{L}_{\text{sci}}(s, \overrightarrow{\theta_s}, \vec{\theta}_b) = \text{Poisson}(N_{\text{sci}} \,|\, \mu_b(\vec{\theta}_b) + \mu_s(s, \overrightarrow{\theta_s}, \vec{\theta}_b)) \times \text{Poisson}(N_{\text{cal}} \,|\, \alpha \times \mu_b(\vec{\theta}_b))$$

Binned Likelihood

- With more than ~ 5 events in each bin, you can use computationally efficient methods to compute test statistic distributions
- Eases visualisation and goodness-of-fit
- And simpler to share results
- Minimal sensitivity loss if the bin width is small compared to the detector resolution

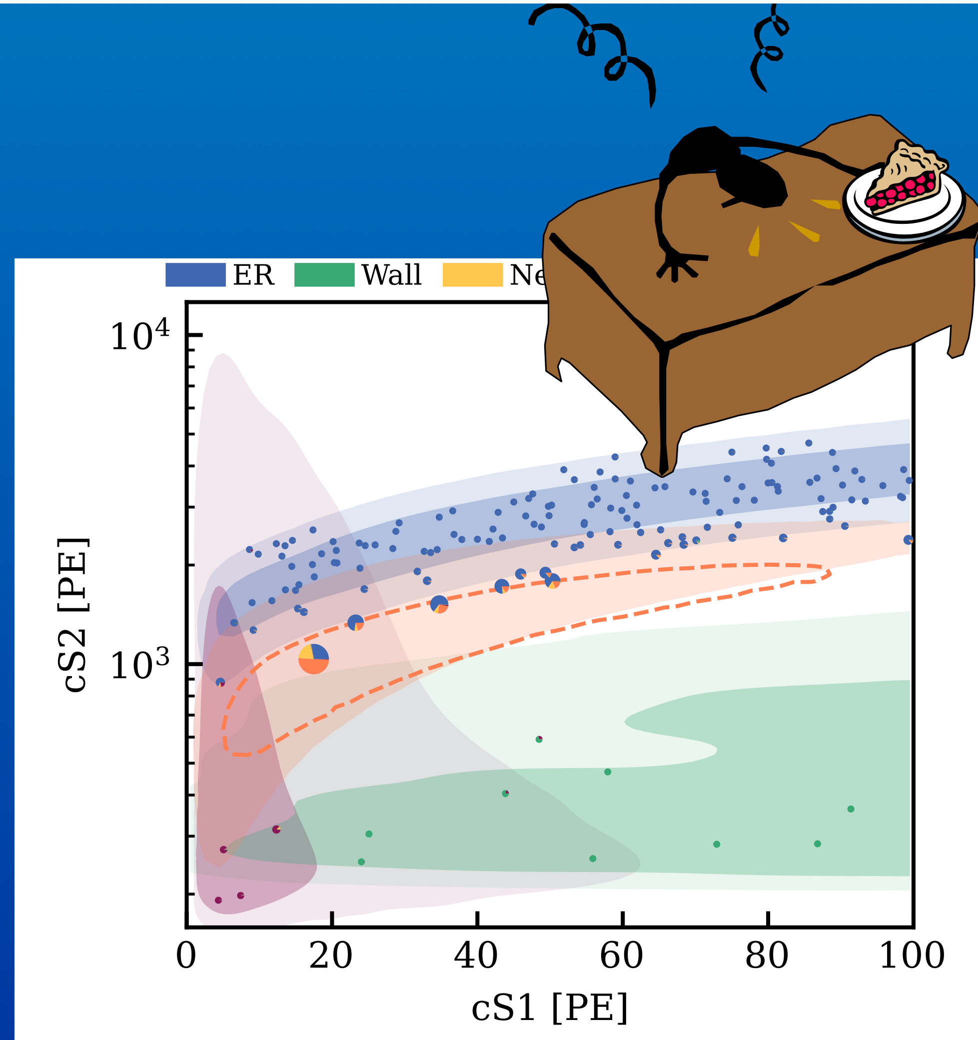


PandaX ionisation-only search, https://arxiv.org/abs/2212.10067

$$\mathscr{L}_{\text{sci}}(s, \overrightarrow{\theta_s}, \vec{\theta}_b) = \prod_{i=1}^{N_s} \left[\text{Poisson}(N_i \,|\, \mu_{b,i}(\vec{\theta}_b) + \mu_{s,i}(s, \overrightarrow{\theta_s}, \vec{\theta}_b)) \right]$$

Unbinned (extended) likelihood

- If the events are too few to fill bins, the unbinned likelihood promises the best performance
- Might still have to rely on binned methods for goodness-of-fit
- if you rely on Monte Carlo methods to generate distributions, that can require a lot of statistics and be harder to validate

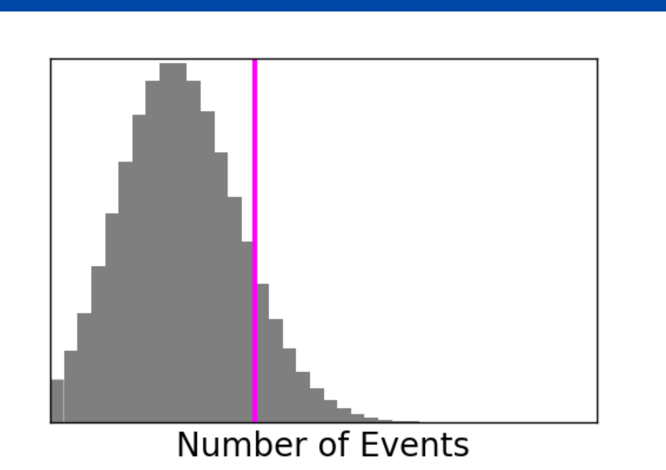


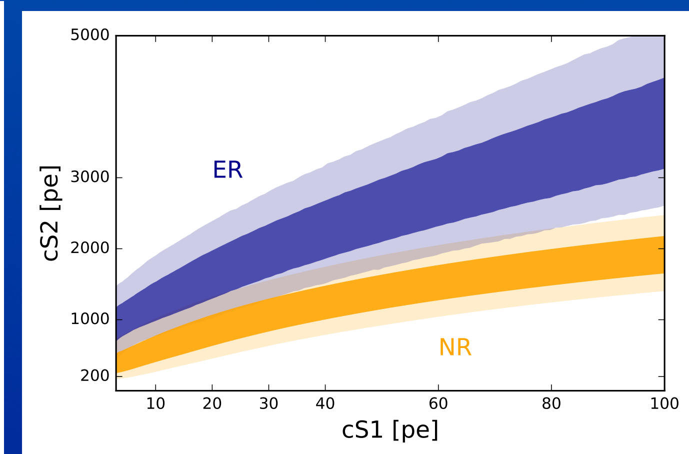
XENONnT first WIMP search

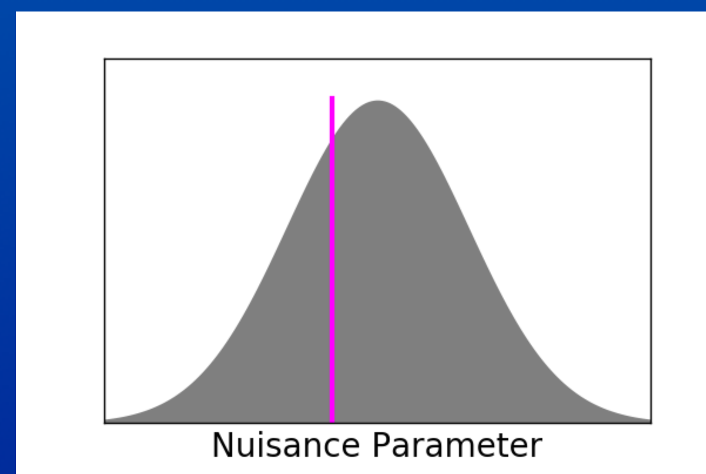
$$\mathscr{L}_{\mathrm{sci}}(s, \overrightarrow{\theta_s}, \vec{\theta}_b) = \mathrm{Poisson}(N_{\mathrm{sci}} \,|\, \mu_b(\vec{\theta}_b) + \mu_s(s, \overrightarrow{\theta_s}, \vec{\theta}_b)) \times \prod_{i=1}^{N_s} \left[\frac{\mu_s}{\mu_s + \mu_b} f_s(\vec{x}_i \,|\, s, \overrightarrow{\theta_s}, \vec{\theta}_b) + \frac{\mu_b}{\mu_s + \mu_b} f_b(\vec{x}_i \,|\, \vec{\theta}_b) \right]$$

Likelihoods can be composed

$$\mathscr{L}(s,\overrightarrow{\theta_s},\vec{\theta}_b)_{\textbf{Science run}} = \mathscr{L}_{\text{sci}}(s,\overrightarrow{\theta_s},\vec{\theta}_b)\times\mathscr{L}_{\text{cal}}(\vec{\theta}_b)\times\mathscr{L}_{\text{anc}}(\vec{\theta}_b)$$





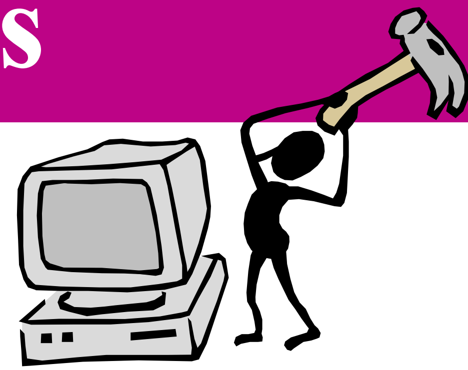


$$\mathscr{L}(s,\overrightarrow{\theta_s},\vec{\theta}_b)_{\textbf{tot}} = \mathscr{L}(s,\overrightarrow{\theta_s},\vec{\theta}_b)_{\textbf{tot}}\times\mathscr{L}(s,\overrightarrow{\theta_s},\vec{\theta}_b)_{\textbf{tot}}\times\mathscr{L}_{\textbf{shared}}(\boldsymbol{\theta})$$

Asymptotic Distributions

SoUP 24

A massive shortcut if you're careful/lucky



- The log-likelihood for a number of gaussian-distributed numbers has the same form as the χ^2-formula (Wilks' theorem)
- It turns out that if a set of conditions that are quite often™ fulfilled, the distribution of the likelihood ratio converges to a χ^2-distribution with some number of free parameters
- This can massively simplify your computations, and so it is worth to look through in detail

$$q(s) = -2 \cdot \log\left(\frac{\mathscr{L}(s, \hat{\hat{s}})}{\mathscr{L}(\hat{s}, \hat{s})}\right)$$

Necessary conditions for Wilks' theorem

ASYMPTOTIC: Sufficient data is observed.

INTERIOR: Only values of μ and θ which are far from the boundaries of their parameter space are admitted.

IDENTIFIABLE: Different values of the parameters specify distinct models.

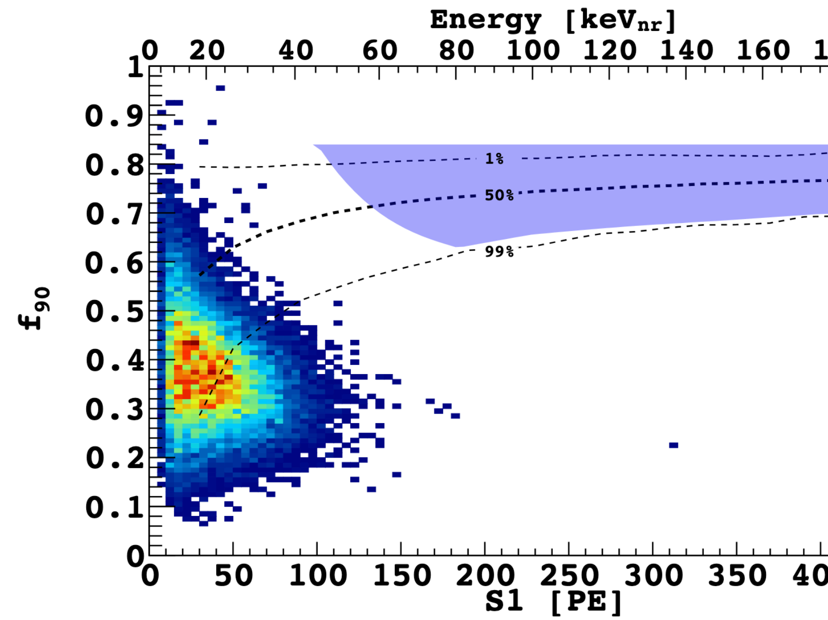
NESTED: H_0 is a limiting case of H_1, e.g. with some parameter fixed to a sub-range of the entire parameter space.

CORRECT: The true model is specified either under H_0 or under H_1.

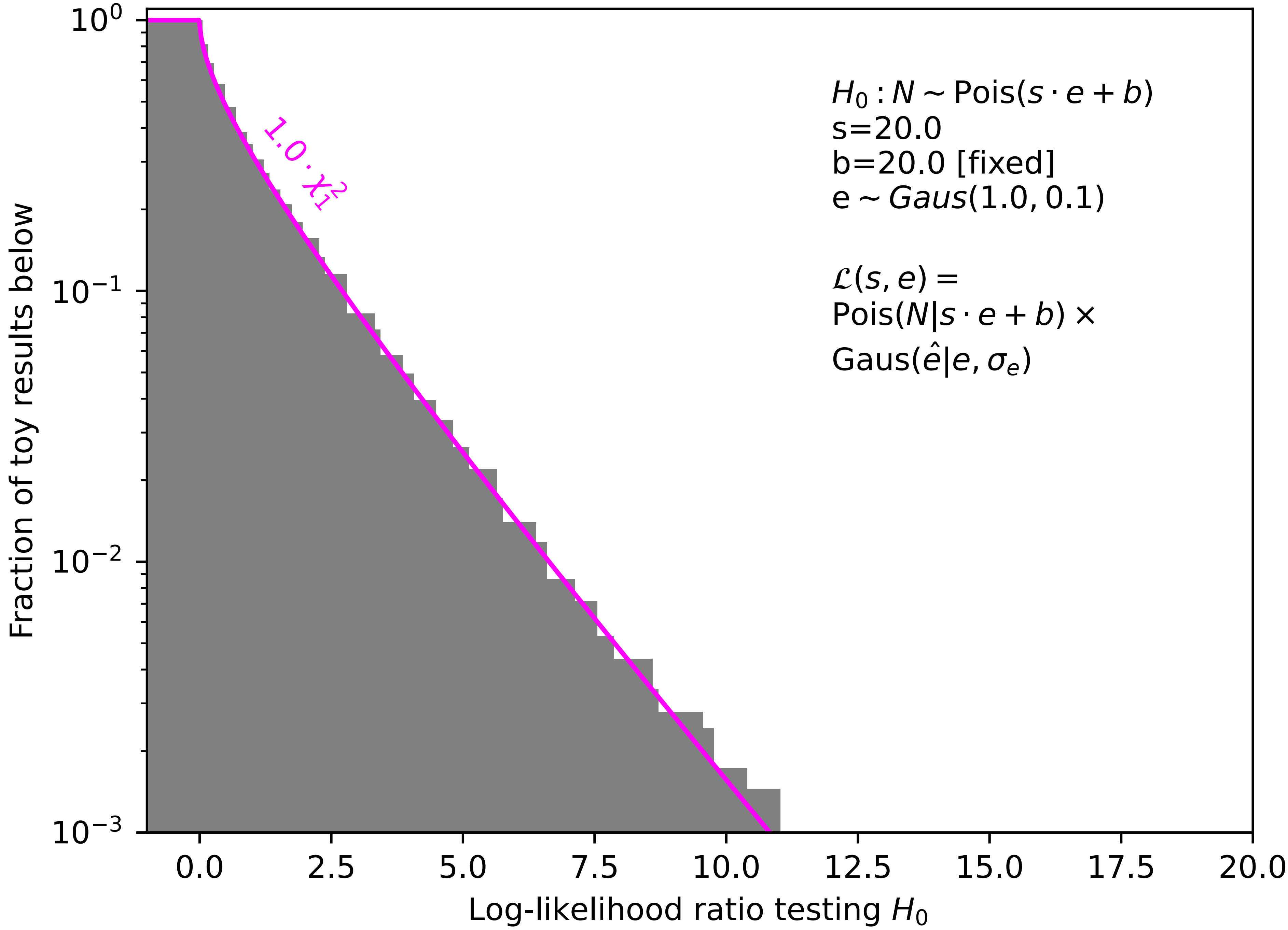
https://arxiv.org/abs/1911.10237

Asymptotic Distributions

- As our example: the profile log-likelihood ratio test for a counting experiment with a known background but uncertain efficiency
- Parameters:
  - Signal s
  - efficiency e
- Fixed, known parameters:
  - Background expectation b
  - efficiency uncertainty σ_e
- Data:
  - Number of events N
  - efficiency estimate e_{meas}



$$(L)(s, e) = \text{Pois}(N \,|\, s \cdot e + b) \times \text{Gaus}(e_{\text{meas}} \,|\, e, \sigma_e)$$


$H_0 : N \sim \text{Pois}(s \cdot e + b)$
s=20.0
b=20.0 [fixed]
e $\sim Gaus(1.0, 0.1)$
$\mathcal{L}(s, e) =$
$\text{Pois}(N|s \cdot e + b) \times$
$\text{Gaus}(\hat{e}|e, \sigma_e)$
$1.0 \cdot \chi_1^2$
Fraction of toy results below
Log-likelihood ratio testing H_0
10^0
10^{-1}
10^{-2}
10^{-3}
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

Asymptotic Distributions

SoAP 24

What does "sufficiently data" mean?

- Wilks' theorem holds in the asymptotic case of infinite data, but convergence can often be quick:
    - Poisson counting with more than ca. 10 events
    - Gaussian measurements
- However, if you have an unbinned likelihood, the important consideration is *signal-like* background events— for example seen with lXe TPC searches

Necessary conditions for Wilks' theorem

ASYMPTOTIC: Sufficient data is observed.

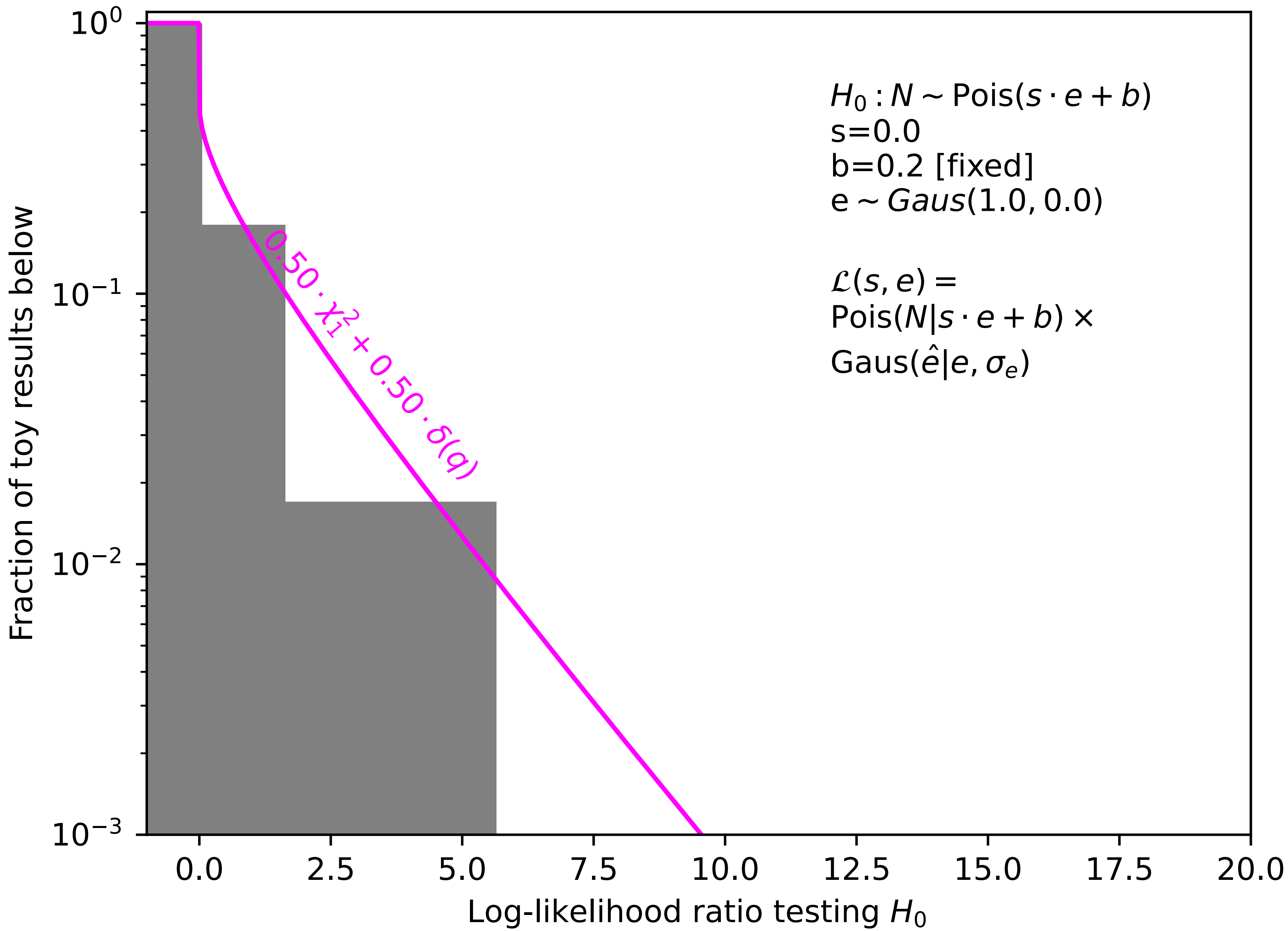
INTERIOR: Only values of μ and θ which are far from the boundaries of their parameter space are admitted.

IDENTIFIABLE: Different values of the parameters specify distinct models.

NESTED: H_0 is a limiting case of H_1, e.g. with some parameter fixed to a sub-range of the entire parameter space.

CORRECT: The true model is specified either under H_0 or under H_1.

https://arxiv.org/abs/1911.10237



$H_0 : N \sim \text{Pois}(s \cdot e + b)$
s=0.0
b=0.2 [fixed]
$e \sim Gaus(1.0, 0.0)$
$\mathcal{L}(s, e) =$
$\text{Pois}(N|s \cdot e + b) \times$
$\text{Gaus}(\hat{e}|e, \sigma_e)$
$0.50 \cdot \chi_1^2 + 0.50 \cdot \delta(q)$
Fraction of toy results below
Log-likelihood ratio testing H_0

Asymptotic Distributions

SoUP 24

What does "interior of the parameter space" mean?

- As a mental shortcut— if under your null or signal hypothesis, parameters sometimes or often goes to a physical boundary, it will not behave asymptotically
- This is very often the case e.g. if you're looking for a signal with expectation value ≥ 0
- If you are testing the hypothesis that the model that has the parameter *at the boundary*— for example that the signal is 0, you may be able to use *Chernoff's theorem* if all other conditions are met

Necessary conditions for Wilks' theorem

ASYMPTOTIC: Sufficient data is observed.

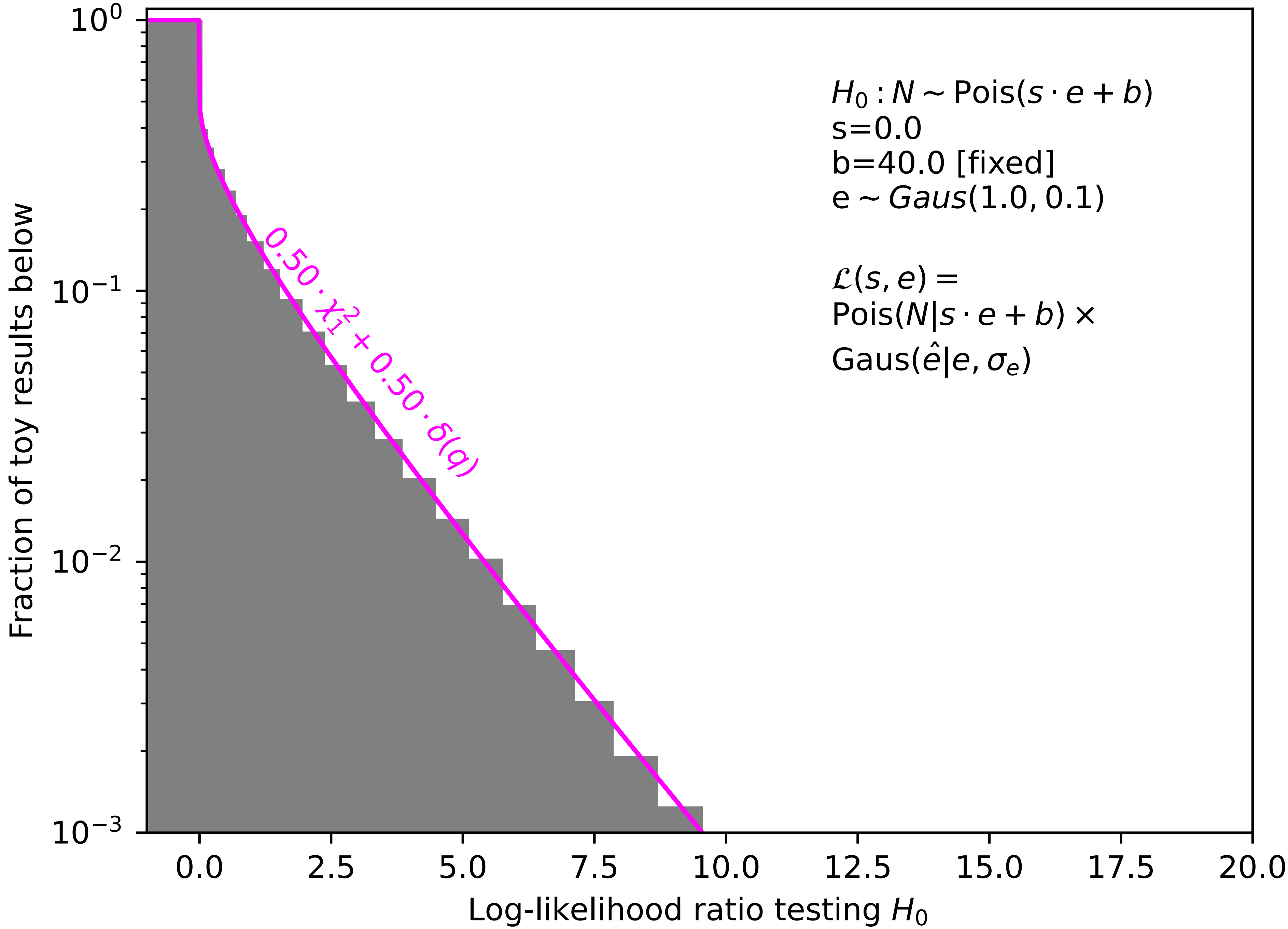
INTERIOR: Only values of μ and θ which are far from the boundaries of their parameter space are admitted.

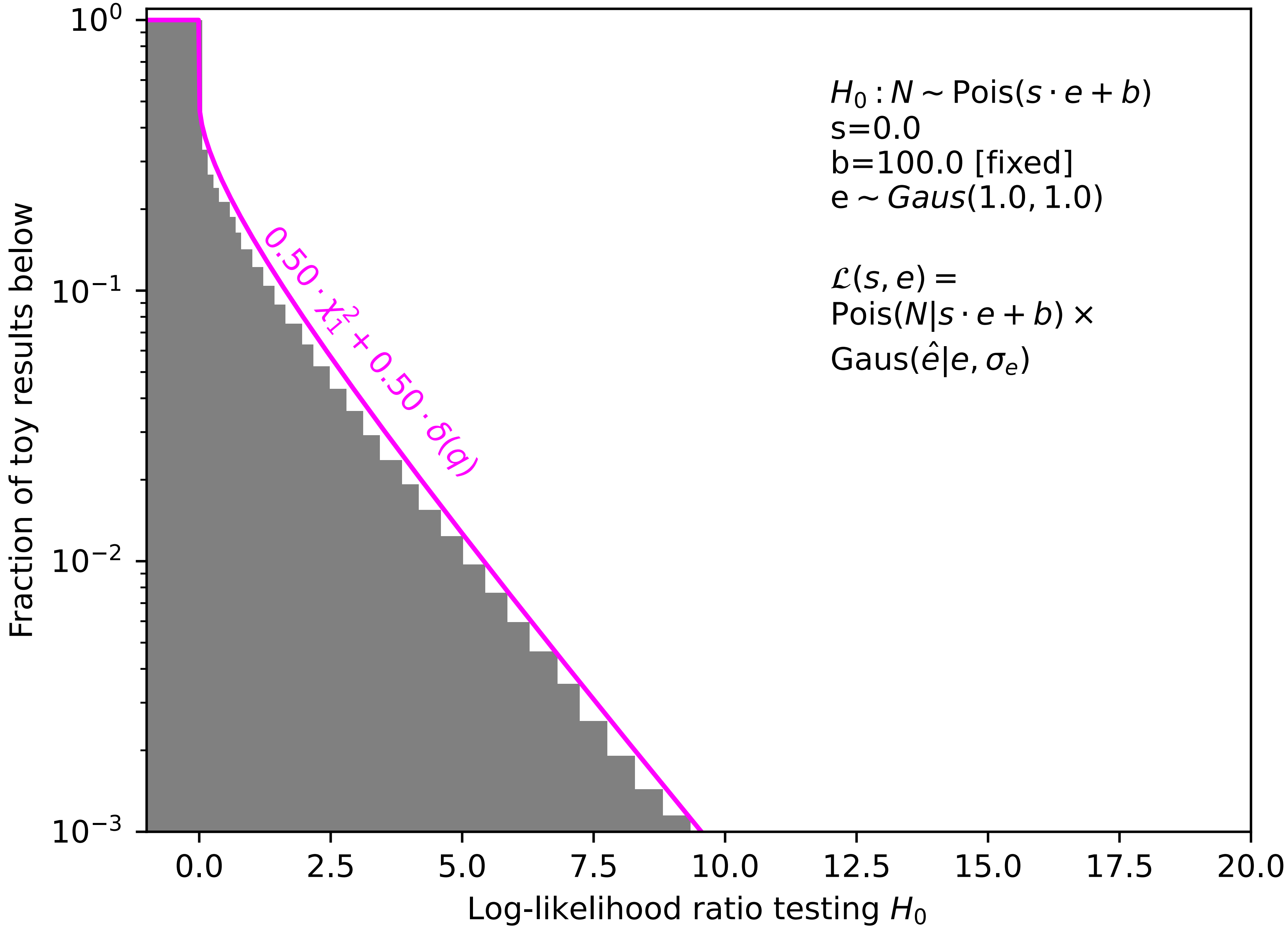
IDENTIFIABLE: Different values of the parameters specify distinct models.

NESTED: H_0 is a limiting case of H_1, e.g. with some parameter fixed to a sub-range of the entire parameter space.

CORRECT: The true model is specified either under H_0 or under H_1.

$$f(q) \overset{\text{Chernoff}}{\approx} \frac{1}{2}\chi^2_{DOF=1} + \frac{1}{2}\delta(\hat{\mu})$$


$H_0: N \sim \mathrm{Pois}(s \cdot e + b)$
s=0.0
b=40.0 [fixed]
$\mathrm{e} \sim Gaus(1.0, 0.1)$
$\mathcal{L}(s, e) =$
$\mathrm{Pois}(N|s \cdot e + b) \times$
$\mathrm{Gaus}(\hat{e}|e, \sigma_e)$
$0.50 \cdot \chi_1^2 + 0.50 \cdot \delta(q)$
Fraction of toy results below
Log-likelihood ratio testing H_0
10^0
10^{-1}
10^{-2}
10^{-3}
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0



$H_0 : N \sim \mathrm{Pois}(s \cdot e + b)$
s=0.0
b=100.0 [fixed]
$e \sim Gaus(1.0, 1.0)$
$\mathcal{L}(s, e) =$
$\mathrm{Pois}(N|s \cdot e + b) \times$
$\mathrm{Gaus}(\hat{e}|e, \sigma_e)$
$0.50 \cdot \chi^2_1 + 0.50 \cdot \delta(q)$
Fraction of toy results below
Log-likelihood ratio testing H_0

Asymptotic Distributions

SoUP 24

What does it mean for parameters to be "identifiable"?

- If the model is degenerate for some parameter, the asymptotic approximation will not hold
- This is quite common in physics! When the signal strength is 0, the model does not depend on any other signal parameter
- This is another way of looking at the look-elsewhere effect, which we'll look at later

Necessary conditions for Wilks' theorem

ASYMPTOTIC: Sufficient data is observed.

INTERIOR: Only values of μ and θ which are far from the boundaries of their parameter space are admitted.

IDENTIFIABLE: Different values of the parameters specify distinct models.

NESTED: H_0 is a limiting case of H_1, e.g. with some parameter fixed to a sub-range of the entire parameter space.

CORRECT: The true model is specified either under H_0 or under H_1.

Asymptotic Distributions

What does it mean for models to be "nested"?

- If the model tested is not a limit of the general hypothesis
  - Such as when testing between two disparate models
  - Or if your theory features a non-zero fixed signal you wish to test against the no-signal hypothesis
- You can always linearly add the two hypotheses' models together with a new parameter, but then you introduce Non-identifiability at the boundary!

Necessary conditions for Wilks' theorem

ASYMPTOTIC: Sufficient data is observed.

INTERIOR: Only values of μ and θ which are far from the boundaries of their parameter space are admitted.

IDENTIFIABLE: Different values of the parameters specify distinct models.

NESTED: H_0 is a limiting case of H_1, e.g. with some parameter fixed to a sub-range of the entire parameter space.

CORRECT: The true model is specified either under H_0 or under H_1.

https://arxiv.org/abs/1911.10237

Asymptotic Distributions

SoUP 24

The models still need to be correct :(

- All our inference results are reliant on the true model being somewhere in our model space!
- However, we should be cognisant that this is never guaranteed
- If you have a mismodelling you are concerned about, you should test how much it can affect your results— you might well find that your method is robust to it, or you can add model uncertainties to represent this
- Another way to increase robustness is to make your model simpler— a counting experiment makes fewer assumptions on the energy spectrum than if you include the energy information

Necessary conditions for Wilks' theorem

ASYMPTOTIC: Sufficient data is observed.

INTERIOR: Only values of μ and θ which are far from the boundaries of their parameter space are admitted.

IDENTIFIABLE: Different values of the parameters specify distinct models.

NESTED: H_0 is a limiting case of H_1, e.g. with some parameter fixed to a sub-range of the entire parameter space.

CORRECT: The true model is specified either under H_0 or under H_1.

https://arxiv.org/abs/1911.10237

Asymptotic Distributions

SoUP 24

Examples of when they may be used

$$q_{\text{discovery}} = \begin{cases} q(0) \text{ if } 0 \leq \hat{\mu} \\ 0 \text{ else} \end{cases}$$

- Any gaussian-distributed measurements
  - Including histograms with high bin counts

$$q_{\text{upper limit}}(\mu) = \begin{cases} q(\mu) \text{ if } \hat{\mu} \leq \mu \\ 0 \text{ else} \end{cases}$$

- unbinned likelihoods with significant signal-like backgrounds

$$q(\mu)_{\text{unified}} = \begin{cases} -2 \cdot \frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}})}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})} \text{ if } 0 \leq \hat{\mu} \\ -2 \cdot \frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}})}{\mathcal{L}(0, \hat{\boldsymbol{\theta}}_{\mu=0})} \text{ else} \end{cases}$$

- Most common extra consideration is taking care of the parameter boundaries
- The below paper presents some cases:

https://arxiv.org/abs/1007.1727

Note that these three can be seen as the same test statistic if you always restrict $\hat{\mu}, \mu$ to be positive!

The Look Elsewhere Effect

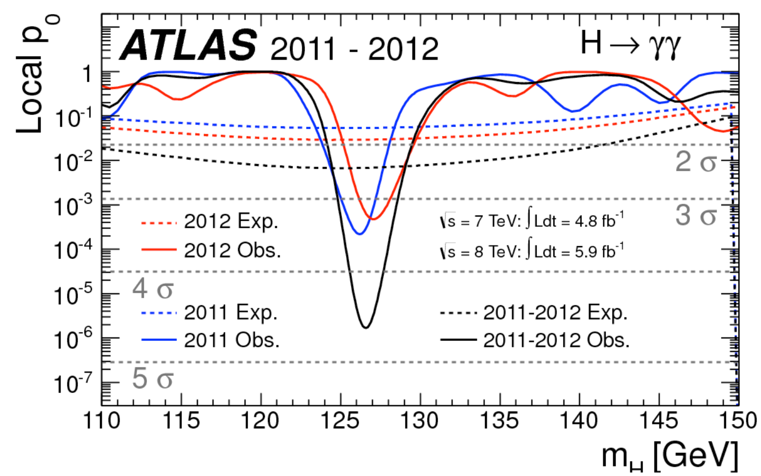
AKA trial factor AKA non-identifiable signal parameters

- Your probability to roll 6 on a dice increases the more dice you get to roll

- Similarly, if your experiment tests several signals, they will increase their chance to see unusual effects just by chance

- Separate between "local" significance— the probability that one single signal model tests fluctuates to some significance

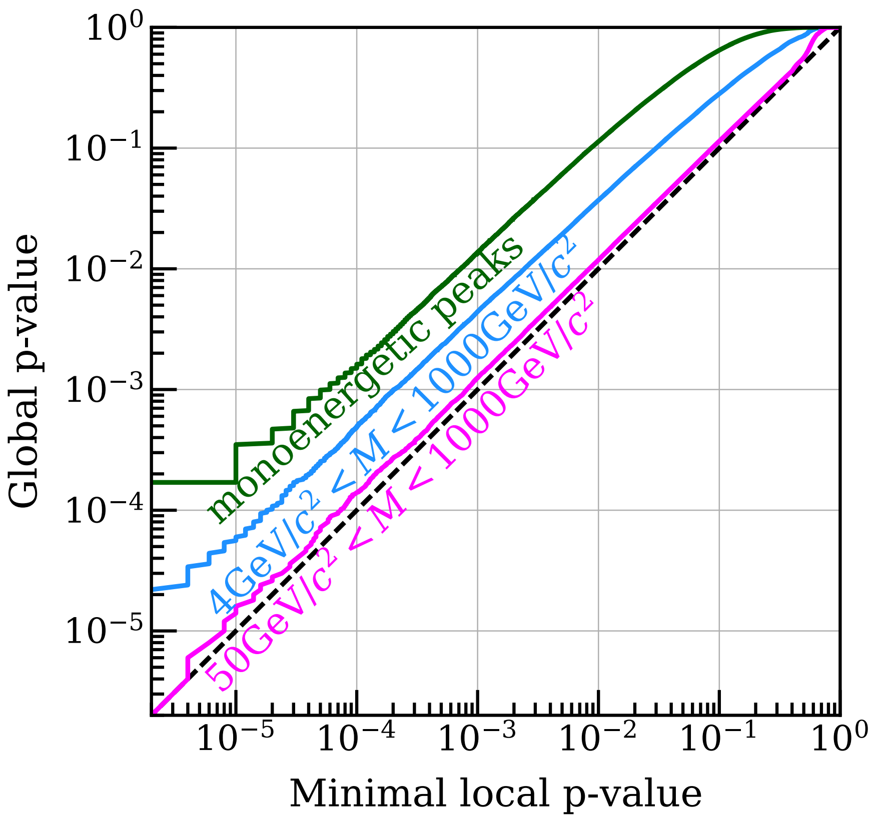
- and "global" significance— the probability that *any* test fluctuates to that extent

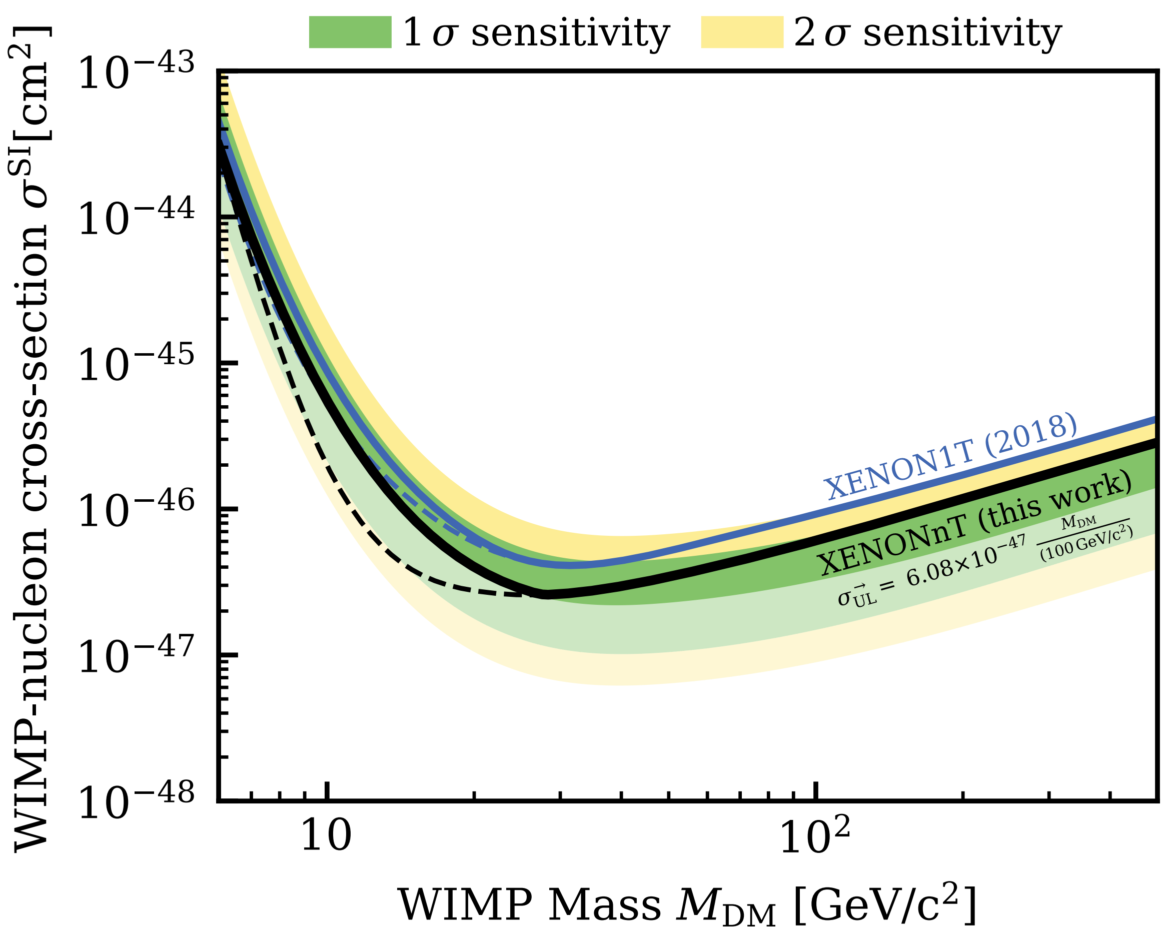


The Look Elsewhere Effect

SoUP 24

The trial factor might sometimes be rather small:

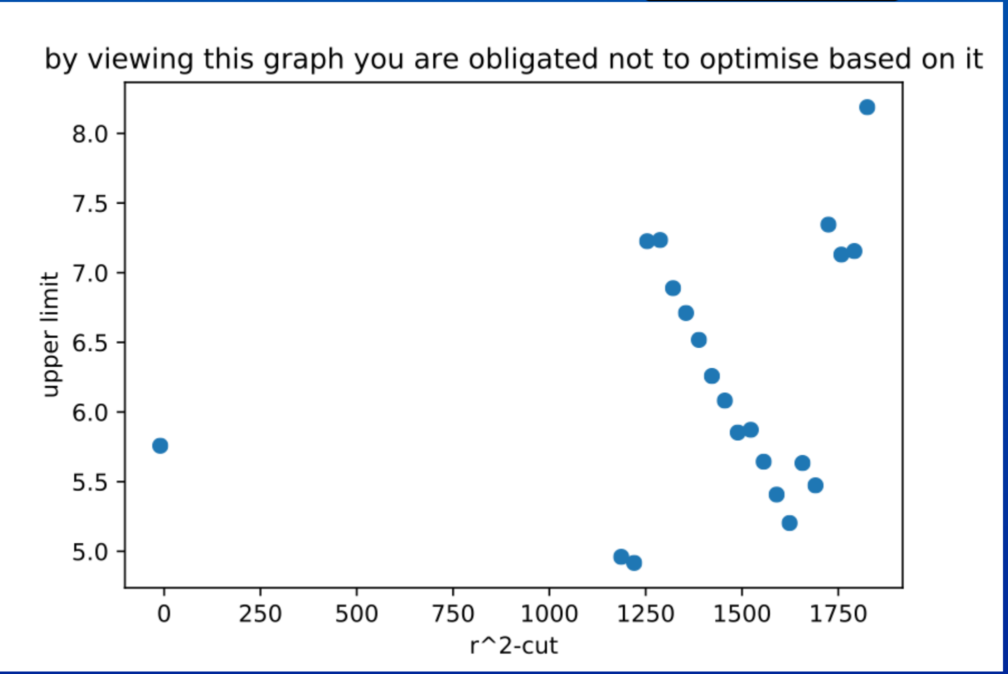




Experimenter bias is a danger with few events

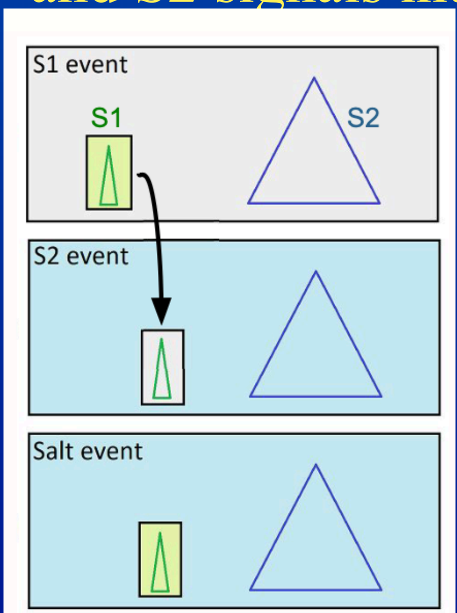
- With few events the effect can be drastic if you chance something in your analysis—the plot shows the 60% change in limit available to you between the best post-unblinding and the worst post-unblinding radial cut.
  - This is a necessary consequence of making your analysis sensitive to few events!
- Further, with only some hundreds of events, and many variables, every event may well be an outlier in some space

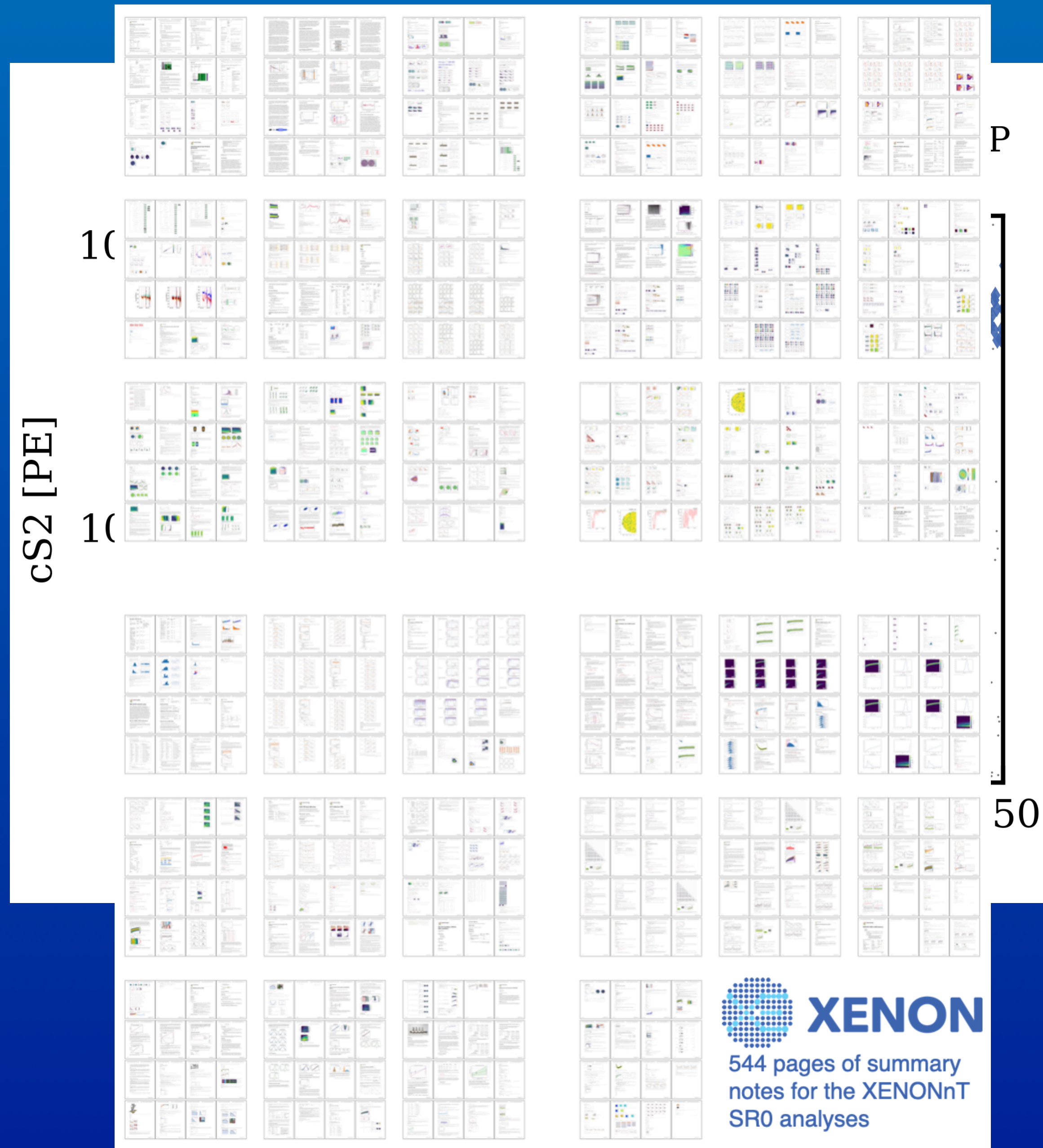
Homeopathic poison — the fewer events the greater danger



Experimenter bias is a danger with few events

- The most common experimenter bias mitigation method is "blinding"— not showing the signal-like region of parameter space until the analysis has been frozen
- LUX developed a "salting" procedure where synthetic signals were made by stitching together genuine S1 and S2 signals into full events in the data





Tyler Anderson "Salting as a Bias Mitigation Technique in LZ", presentation at LIDINE 2021

For today

- Example analyses: we saw how experiments compose analytical and other models to make their full statistical model
- Profile Likelihood: we discussed minimising nuisance parameters only, if we wish to test some hypothesis
- Asymptotic distributions: How useful they are, and the common failures we encounter
- Look-Elsewhere effect: one of these effects

Hands-on session: profile likelihood, comparison with the asymptotic:

SOUP 2024 Exercise set 2: Likelihood optimisation, confidence intervals, asymptoticity checks

Welcome to the SOUP exercise set!

```
In [1]:
# Import modules with the tools we need:
import scipy.stats as sps
from scipy.optimize import minimize
import matplotlib.pyplot as plt
import numpy as np
```

0: Optimisation

scipy.optimize.minimize (or iminuit if you prefer) will be our bread-and-butter
The key trick is to remember how to set up the function to be used:

```
In [2]:
def function_scipy(x, a=0.5, b=1):
    #if 0<a, this function has a global minimum at 0,0, but if a is close to 0, this minimum will have a broad nea
    #note that only parameters inside the vector x are minimized for-- the rest are fixed parameters.
    return x[0]**2 + b*(x[1]-a*x[0]**2)**4
```

```
In [3]:
#minimize using scipy:
scipy_result = minimize(function_scipy, [3, 9.2], args=(0.5,1)) #notice that you have to provide a guess for "x"
#and that you can pass further args like a,b to the minimizer
print(scipy_result)
```

```
      fun: 1.2859447158387182e-08
 hess_inv: array([[3.80897491e-01, 1.25210924e-02],
       [1.25210924e-02, 3.07946843e+02]])
      jac: array([ 9.02614194e-07, -4.83026339e-06])
  message: 'Optimization terminated successfully.'
     nfev: 102
      nit: 24
     njev: 34
   status: 0
  success: True
        x: array([ 4.43855426e-07, -1.06488795e-02])
```

Iminuit

Iminuit is heavily used in particle physics, and has some nice extra functions

```
In [4]:
def function_minuit(x,y, a=0.5, b=1):
    return function_scipy([x,y],a, b)
```

```
In [5]:
try:
    import iminuit
    mobj = iminuit.Minuit(function_minuit, x=3,y=9.2, a=0.5, b=1)
    mobj.fixed["a"] = True #here, we can set explicitly parameters to be fixed (and free them later!)
    mobj.fixed["b"] = True #here, we can set explicitly parameters to be fixed (and free them later!)
    minuit_result = mobj.migrad() #call the minimizer routine (called Migrad)
    print(minuit_result)
except:
    minuit_result = None
    print("Iminuit is not installed, install it or use scipy.optimize")
```

```
┌──────────────────────────────────────────────────────────────────┐
│                              Migrad                              │
├──────────────────────────────┬───────────────────────────────────┤
│ FCN = 1.475e-06              │              Nfcn = 161           │
│ EDM = 1.15e-06 (Goal: 0.0002)│                                   │
├──────────────────────────────┼───────────────────────────────────┤
│        Valid Minimum         │     No Parameters at limit        │
├──────────────────────────────┼───────────────────────────────────┤
│ Below EDM threshold (goal x 10) │       Below call limit         │
├───────────────┬──────────────┼──────────┬───────────┬────────────┤
│  Covariance   │   Hesse ok   │ Accurate │ Pos. def. │ Not forced │
└───────────────┴──────────────┴──────────┴───────────┴────────────┘
```

	Name	Value	Hesse Err	Minos Err-	Minos Err+	Limit-	Limit+	Fixed
0	x	-0	1					
1	y	-0	13					
2	a	0.500	0.005					yes
3	b	1.00	0.01					yes

	x	y	a	b